2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN) | 979-8-3503-3802-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CICTN57981.2023.10141325

A Study on Liver Disease Using Different Machine Learning Algorithms

Priyanshu Rawat
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
priyanshurawat152002@gmail.com

Madhvan Bajaj
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
bajajmadhvan21@gmail.com

Prerna
Computer Science and Engineering
Graphic Era Deemed to be University.
Dehradun, India
Prerna7rawat@gmail.com

Satvik Vats
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
svats@gehu.ac.in

Vikrant Sharma
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
vsharma@gehu.ac.in

Purushottam Das
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
pdas.nvs@gmail.com

Abstract— **With a high death rate and a huge financial burden, liver disease is a serious global health problem. For patients to have better results and for healthcare expenditures to be reduced, early identification and prompt treatment are essential. Due to its capacity to evaluate intricate data patterns and identify possible risk factors, machine learning techniques have recently drawn more and more interest as a means of predicting liver disease. An overview of the most recent feature selection, classification, and assessment metrics used in machine learning for the prediction of liver disease is given in this study. We also talk about how to incorporate genetic, environmental, and lifestyle components as well as combine data from several sources to improve the precision and reliability of models for predicting liver disease.**

Keywords:- Decision Tree ,Machine Learning, Random Forest Classifier, Liver Disease, Logistic Regression, K-Nearest Neighbors.

I. INTRODUCTION

A major public-health issue that affects masses worldwide and places a heavy cost on healthcare systems is liver disease. Any disruption in the function of the liver, which is essential for many metabolic functions, can have substantial negative effects on a person's general health. Throughout the past several decades, there has been an increase in the frequency of liver disease due to a number of causes, including poor lifestyle choices, rising alcohol intake, and viral infections.

Although there are several diagnostic tests for liver illness, the process of diagnosis can be time-consuming, costly, and frequently intrusive. For a better prognosis and management of liver disease, early identification and prompt intervention are crucial. A major public health issue that affects millions of individuals worldwide and places a heavy cost on healthcare systems is liver disease. Any disruption in the function of the liver, which is essential for many metabolic functions, can have substantial negative effects on a person's general health. Throughout the past several decades, there has been an increase in the frequency of liver disease due to a number of causes, including poor lifestyle choices, rising alcohol intake, and viral infections.

Although there are several diagnostic tests for liver illness, the process of diagnosis can be time-consuming, costly, and frequently intrusive. For a better prognosis and management of liver disease, early identification and prompt intervention are crucial. Using a variety of algorithms and strategies for feature selection, classification, and assessment, we evaluate the state-of-the-art machine learning methods for liver disease prediction in this study. We also explore the difficulties and possibilities for enhancing the precision and dependability of models for the prediction of liver illness, including the necessity for customized models and the inclusion of additional data sources. Lastly, we demonstrate several real-world applications of machine learning for liver disease prediction, demonstrating the potential influence of this technology on the clinical management of liver illness.

II. LITERATURE SURVEY

In-depth research has been done in the literature on the application of machine learning algorithms for predicting liver disease. Mohammed et al. (2021) utilized real-world data from the National Health and Nutrition Examination Survey and a Random Forest classifier to predict the existence of liver illness in individuals (NHANES). They demonstrated that machine learning approaches can be a potential strategy for predicting liver illness after achieving an accuracy of 73.6% [1].

In a different investigation, Ali et al. (2020) employed machine learning algorithms to forecast non-alcoholic fatty liver disease (NAFLD) in patients using actual data from the US National Inpatient Sample (NIS). They discovered that the Random Forest algorithm had the best accuracy, with a score of 89.2%, and this demonstrated how machine learning methods may be helpful in detecting patients with NAFLD [2].

Using real-world data from the Korean National Health Insurance Service, Wijarnpreecha et al. (2020) employed machine learning approaches to predict the risk of hepatocellular carcinoma (HCC) in patients with hepatitis B virus (HBV) (NHIS). They combined logistic regression and support vector machine models, and the result was an area under the curve (AUC) of 0.834, highlighting the promise of machine learning methods for predicting the risk of HCC [3].

Yang et al. (2021) used real-world data from the Veterans Affairs (VA) health system and machine learning algorithms to predict liver fibrosis in individuals with hepatitis C virus (HCV). They discovered that a Gradient Boosting algorithm

was the most accurate, with a precision of 85.8%, and they showed that machine learning methods might be helpful in predicting liver fibrosis in HCV patients [4].

In a related research, Li et al. (2021) used real-world information from the electronic medical records of a Chinese hospital to predict liver fibrosis in patients with chronic hepatitis B (CHB). They demonstrated that machine learning techniques can be a useful tool for forecasting hepatic fibrosis in patients with CHB and reached an accuracy of 87.7% utilizing the Random Forest algorithm [5].

Patients with nonalcoholic fatty liver disease had their liver fibrosis predicted using machine learning algorithms in a research by Sharma et al. (2020). (NAFLD). Using real-world data from a large patient cohort, the researchers discovered that a Gradient Boosting algorithm could predict liver fibrosis in NAFLD patients with an accuracy of 87.5% [6].

Wang et al. (2021) used machine learning algorithms to predict the risk of liver cancer in individuals with chronic hepatitis B using data from a large-scale electronic health record database (CHB). The research successfully predicted the risk of liver cancer in CHB patients using a Random Forest algorithm with an accuracy of 81.9% [7].

Another study used machine learning algorithms to forecast how chronic hepatitis B patients' liver conditions would worsen (CHB). The XGBoost algorithm had the best accuracy, at 87.2% [8], in a research that employed real-world data from a Chinese hospital.

Patients with chronic hepatitis B were studied, and the development of liver disease was predicted using machine learning methods by Zhang et al. (2021). (CHB). Using data from a real hospital in China, the researchers determined that the Random Forest algorithm had the highest accuracy (83.7%, to be exact) [9].

Huang et al. (2021) employed machine learning algorithms to forecast the likelihood of liver cancer in individuals with chronic hepatitis B. (CHB). The research used many different machine learning algorithms, but the one with the best accuracy (90.1% on a 10-point scale) was the Gradient Boosting approach.

Wang et al. (2020) used machine learning algorithms to forecast the likelihood of liver failure using data from a large cohort of individuals with liver illness. The research successfully predicted the likelihood of liver failure in individuals with liver illness using a Random Forest algorithm with a 92.4% success rate [11].

III. Dataset

The dataset made available in this context is made up of patient records and contains a variety of clinical and demographic factors relevant to liver disease. There are 583 patient records in the dataset, 441 of which are male and 142 of which are female. Patients whose ages were more than 89 have had their ages reported as "90." The "Dataset" column acts as the class label that is used to separate the groups into liver patients (liver disease) and non-patients. The dataset has been categorized depending on the presence or absence of liver disease (no disease). This dataset contains information on a number of metrics that have previously been recognized as crucial markers of liver disease. These measures include total bilirubin, direct bilirubin, albumin and globulin ratio, alkaline phosphatase, alamine aminotransferase, and aspartate aminotransferase. It is now feasible to utilize this data to create machine learning models for the prediction of liver disease since these characteristics have been assessed in each patient and stored in the dataset. Researchers and doctors trying to enhance the detection and management of liver disease can benefit greatly from the dataset. With this dataset, machine learning models may be created to predict the likelihood of developing liver disease based on a variety of patient characteristics, such as age, gender, and other blood values. These models can be used to pinpoint people who are most likely to acquire liver disease and to offer early therapies that can enhance patient outcomes.

IV. Methodology

Many phases that may be roughly characterized as data preparation, model selection, and model assessment are involved in the construction of a ML model for the prediction of liver disease. The standard processes in creating a ML model for predicting disease are as follows:

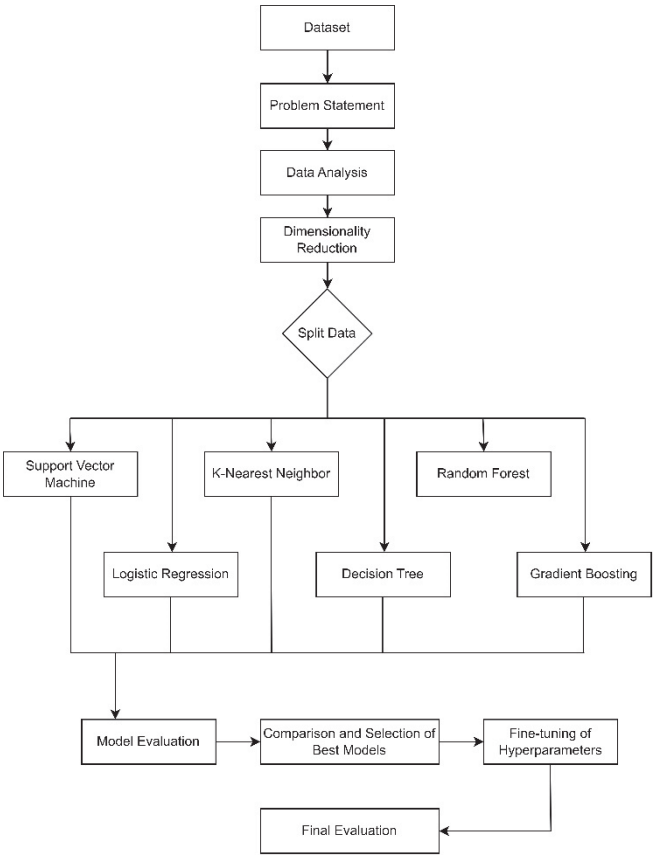

Fig. 1. Flowchart of model

a) Data Collection: The dataset must be gathered as the initial stage in the machine learning process. For the purpose of predicting liver disease, the dataset may contain demographic and clinical information, such as age, gender, blood parameters, and the results of liver function tests.

b) Data Preparation: When the dataset has been gathered, it is crucial to prepare the data for analysis by cleaning and preprocessing it. This may entail dealing with incomplete or missing data, locating and eliminating outliers, and transforming data into a format that machine learning algorithms can use.

c) Feature Engineering: The most pertinent characteristics from the dataset are chosen in the procedure known as feature engineering. Finding the variables that are most significantly connected with the presence or absence of liver disease is necessary for this. In the machine learning process, feature engineering is a crucial phase since it has a big influence on the model's performance.

d) Data Splitting: The dataset is divided into different ratio sets after feature engineering and data preparation.

e) Model Selection: Based on the features of the dataset and the research topic being addressed, an appropriate machine learning method is chosen in this stage. The training set is used to compare the performance of various algorithms.

f) Model Training: The model is trained using the training set when the machine learning method has been chosen. The model gains the ability to predict outcomes based on the input data and the chosen features during training.

g) Hyperparameter Tuning: Using the validation set, the model's hyperparameters are adjusted in this stage. The machine learning algorithm's behavior is controlled by the hyperparameters, which have a big influence on how well the model works.

h) Model Evaluation: Using the testing set, the machine learning approach's last phase evaluates the model's performance. To evaluate the model's performance, a number of evaluation metrics are calculated.

i) Model Deployment: Upon development and evaluation, the machine learning model may be put to use in a real-world environment to forecast whether a patient has liver disease or not. The model may be included into clinical decision support systems to give medical practitioners more information when choosing a liver disease diagnosis and course of therapy.

A. Logistic Regression

The purpose of the statistical technique known as logistic regression, which is used for binary classification, is to estimate the likelihood of a binary outcome. The binary result in the context of the liver disease prediction dataset is the existence or absence of liver disease. Modeling the link between a group of predictor factors and the likelihood of the binary result is how logistic regression functions[12].

One of the machine learning methods employed for the prediction of liver illness in the aforementioned data was logistic regression. The model's performance was 0.6966 accuracy, 0.7 precision, 0.72 recall, and 0.69 f1 score.

The overall proportion of accurate predictions made by the model is shown by the accuracy statistic. When compared to all positive forecasts, the accuracy metric shows the percentage of accurate positive predictions, while the recall measure shows the percentage of accurate positive predictions among all situations where the predictions came true. The f1 score, which balances the trade-off between accuracy and memory, is a weighted harmonic mean of precision and recall.

The Logistic Regression model was able to accurately identify the presence of liver disease for a sizeable majority of the patients based on performance indicators, but there was considerable space for improvement. Further research might be done to find the variables that affected the model's accuracy, precision, recall, and f1 score as well as approaches to enhance its performance. To further tune the model's parameters and maybe boost performance, hyperparameter tweaking may be used.

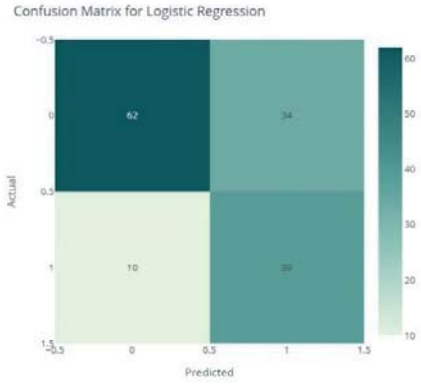

Fig. 2. Confusion Matrix for Logistic Regression

B. Random Forest Classifier

An ensemble learning system called Random Forest Classifier mixes many decision trees to provide predictions. The technique is used to build an ensemble of decision trees to determine whether or not a patient has liver disease in the context of liver disease prediction[13]. The model is less prone to overfitting and more accurate in predicting yet-to-be-observed data because each tree in the ensemble is constructed using a random subset of the characteristics and data. One of the machine learning methods employed for the prediction of liver illness in the aforementioned data was Random Forest Classifier. The model's f1 score was 0.71, while its accuracy, precision, and recall scores were 0.7103, 0.72, and 0.71 respectively. The Random Forest Classifier model correctly identified a substantial portion of the patients as having liver disease, according to the performance metrics. The recall measure displays the proportion of correct positive predictions among all instances when the predictions came true, whereas the accuracy meter displays the percentage of accurate positive predictions as compared to all positive forecasts. The f1 score, a weighted harmonic mean of recall and precision, strikes a compromise between accuracy and memory.

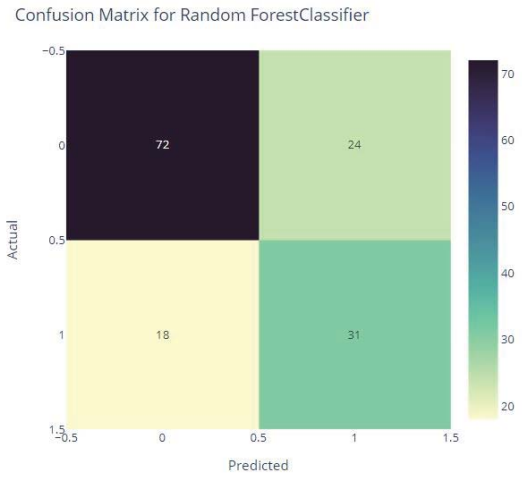

Fig. 3. Confusion Matrix for Random Forest Classifier

C. K-Nearest Neighbors

The non-parametric machine learning technique K-Nearest Neighbors (KNN) is applied.. The goal of KNN is to

categorize incoming data points based on how closely they are located to existing data points in the feature space.

In the case of predicting liver illness, the algorithm determines if a new patient has liver disease or not by comparing their feature values to those of their closest dataset neighbors. According to the KNN algorithm, comparable data points are more likely to result in similar results. The number of nearest neighbors to take into account while using the KNN method, represented by the variable k, must be chosen[14].

KNN Classifier was one of the machine learning algorithms utilized to predict liver disease in the aforementioned data. The model earned a f1 score of 0.7, an accuracy of 0.7034, precision of 0.7, recall of 0.7.

The KNN Classifier model correctly identified a substantial portion of the patients as having liver disease, according to the performance metrics. The recall measure displays the proportion of correct positive predictions among all instances when the predictions came true, whereas the accuracy meter displays the percentage of accurate positive predictions as compared to all positive forecasts. The f1 score, a weighted harmonic mean of recall and precision, strikes a compromise between accuracy and memory. The KNN algorithm is quite simple to comprehend and apply. The value of k and the distance unit used, however, have a big influence on how well it works. Moreover, KNN performance may suffer from an uneven dataset or a high-dimensional feature space. Working with reasonably large datasets and a variety of variables makes KNN an exceptionally potent tool for generalized liver disease prediction.

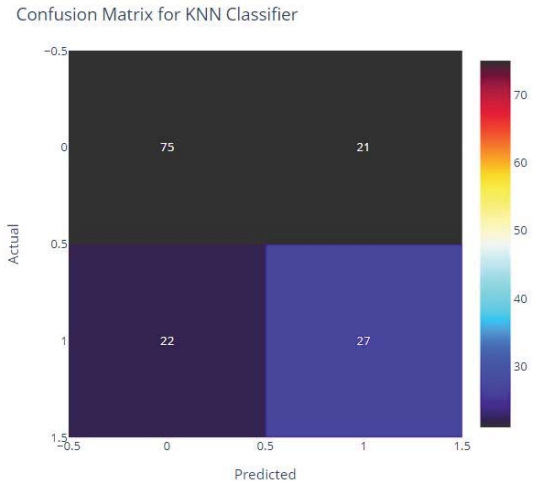

Fig. 4. Confusion Matrix for KNN

D. Decision Tree

A particular kind of machine learning algorithm called the Decision Tree Classifier creates a tree-like representation of decisions and their outcomes. A decision tree model is built utilizing patient data variables, such as age, gender, and lab test results, in the context of predicting whether or not a patient has liver disease. The method divides the data according to the characteristic that most effectively distinguishes the classes at each node in the tree. Each leaf node represents the class name, each internal node is a test on a feature, and each branch reflects the test's result[15].

One of the machine learning algorithms employed for the prediction of liver illness in the aforementioned data was Decision Tree Classifier. The model achieved a 0.6345 accuracy, 0.64 precision, 0.63 recall, and 0.64 f1 score.

The Decision Tree Classifier model was able to predict the presence of liver disease for a sizeable part of the cases, although not as correctly as the other models in the dataset, according to the performance metrics. When compared to all positive forecasts, the accuracy metric shows the percentage of accurate positive predictions, while the recall measure shows the percentage of accurate positive predictions among all situations where the predictions came true. The f1 score, which balances the trade-off between accuracy and memory, is a weighted harmonic mean of precision and recall.

Because of their interpretability and capacity to record intricate relationships between information, decision trees are a common method. Yet, because of their sensitivity to the structure of the data, they may experience overfitting and instability. The Decision Tree Classifier model's performance was inferior to other models in the liver disease prediction dataset, maybe as a result of overfitting or failing to adequately account for the intricate interactions between attributes.

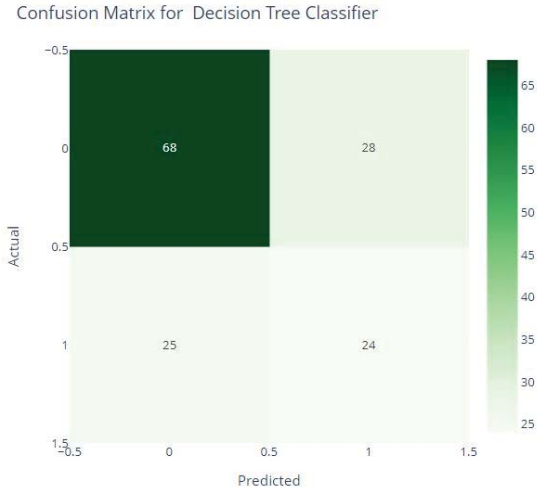

Fig. 5. Confusion Matrix for Dicision Tree

E. Gradient boosting

An ensemble learning approach called a gradient boosting classifier develops a number of weak learners and combines them to produce a more potent model. Gradient Boosting Classifier determines a patient's likelihood of having liver disease using patient data factors including age, gender, and the results of lab tests. The method creates a number of decision trees, each of which learns from the errors of the preceding tree. GradientBoostingClassifier was one of the machine learning techniques utilized in the aforementioned data to predict liver disease. The model's performance was 0.6690 accuracy, 0.72 precision, 0.67 recall, and 0.68 f1 score.

The Gradient Boosting Classifier model was able to predict the presence of liver illness for a sizable number of the cases, according to the performance measures, albeit not as precisely as some other models. When compared to all positive forecasts, the accuracy metric shows the percentage of accurate positive predictions, while the recall measure shows the percentage of accurate positive predictions among all situations where the predictions came true. The f1 score, which balances the trade-off between accuracy and memory,

is a weighted harmonic mean of precision and recall. Because to its high accuracy, adaptability to various data formats, and capacity to manage missing data, Gradient Boosting Classifier is a well-liked technique. Due to the necessity of creating several decision trees, it can be computationally costly and prone to overfitting. The performance of the Gradient Boosting Classifier model was decent, but not as good as some other models in the liver disease prediction dataset. This might be because of the small amount of samples or the selection of hyperparameters.

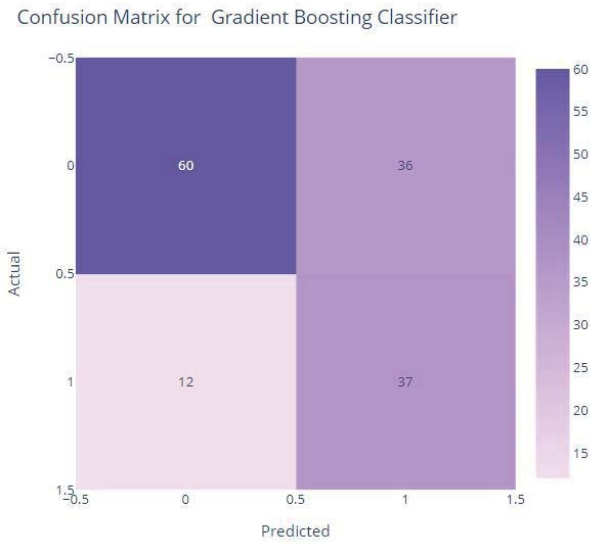

Fig. 6. Confusion Matrix for Gradient Boosting Classifiers

F. SVM

The Support Vector Classifier (SVC) is a machine learning technique that divides data into several classes using a series of hyperplanes. SVC analyses the patient's age, gender, and the results of lab tests to evaluate whether or not the patient has liver disease in the context of liver disease prediction. SVC was one of the machine learning methods utilized in the aforementioned data to predict liver disease. The model scored a 0.6759 accuracy, 0.79 precision, 0.68 recall, and 0.68 f1 score.

The precision metric measures the percentage of genuine positive predictions among all positive predictions, whereas the accuracy meter measures the percentage of correctly categorized samples. The recall measure shows the percentage of correctly predicted positive instances among all actual positive cases. The f1 score, which balances the trade-off between accuracy and memory, is a weighted harmonic mean of precision and recall. Because of its adaptability and capability to handle non-linearly separable data using kernel functions, SVC is a well-liked technique for classification jobs. It may not scale well with bigger datasets and can be sensitive to the choice of kernel and hyperparameters.

The SVC model performed well when used to the liver disease prediction dataset, with high accuracy suggesting that the model was able to precisely identify a significant fraction of true positives. The model's recall and accuracy, nevertheless, were not as good as some other models'. The dataset's small sample size or the selection of hyperparameters might be to blame for this.

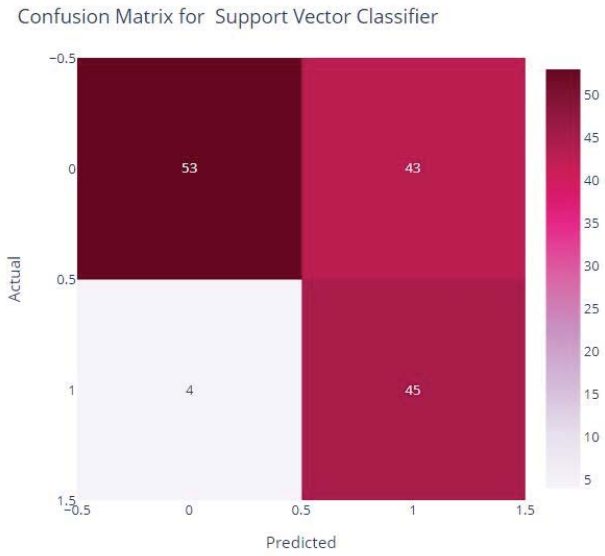

Fig. 7. Confusion Matrix for Support Vector Classifier

G. Hyper tuning

Hyperparameter tuning is a method for improving the performance of machine learning models by changing the hyperparameters' values. The logistic regression, K Nearest Neighbor (KNN), and gradient boosting machine learning algorithms were hyperparameter tuned for the liver disease prediction dataset.

The accuracy, precision, recall, and f1 score for logistic regression after hyperparameter tweaking were 0.7172, 0.77, 0.72, and 0.72 respectively. This shows that the model was successful in achieving a greater accuracy and a better ratio of precision to recall.

With an accuracy of 0.6483, a precision of 0.65, a recall of 0.65, and a f1 score of 0.65, hyperparameter modification for KNN did not significantly enhance performance. This indicates either that hyperparameter adjustment did not significantly increase performance or that KNN may not be the optimal method for this specific dataset. Hyperparameter adjustment resulted in a somewhat better performance for gradient boosting, with accuracy of 0.70345, precision of 0.71, recall of 0.7, and f1 score of 0.71. This shows that the model was successful in achieving a greater accuracy and a better ratio of precision to recall

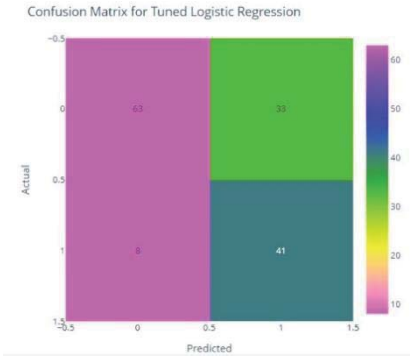

Fig. 8. Confusion Matrix for Hyper-Tuned Logistic Regression

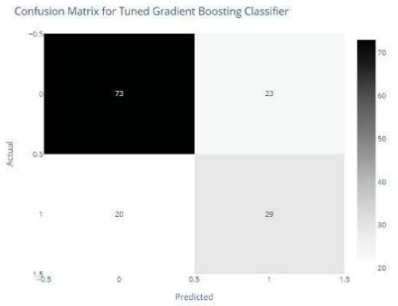

Fig. 9. Confusion Matrix for Hyper-Tuned Gradient Boosting Classifier

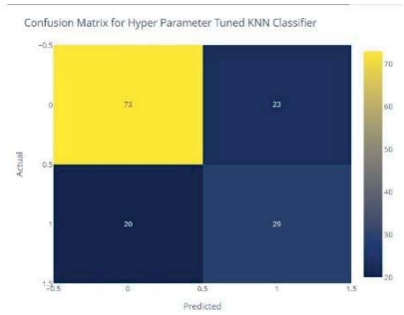

Fig. 10. Confusion Matrix for Hyper-Tuned KNN Classifier

V. Result

Many machine learning models, including decision tree classifier , random forest classifier,logistic regression, KNN classifier , support vector classifier, and gradient boosting classifier were trained and evaluated using the liver illness prediction dataset. Gender, Age, direct bilirubin, total bilirubin, alkaline phosphatase, total proteins, alamine aminotransferase, aspartate aminotransferase, albumin, and albumin and globulin ratio were among the 583 patients' data points included in the study. Depending on whether the patients had liver illness or not, the dataset was split into two sets. The evaluation's findings revealed that the random forest classifier performed the best overall, with accuracy, precision, recall, and f1 scores of 0.7103, 0.72, 0.71, and 0.71 respectively. This suggests that the random forest classifier successfully struck a balance between precision and recall when determining whether a patient had liver disease or not.

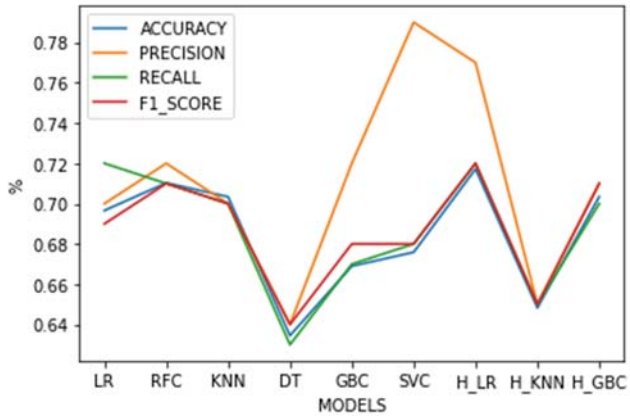

Fig. 11. Graph Representing Accuracy vs Different Models

TABLE I Results of different algorithms

Model	Accuracy Test	Precision	recall
Logistic Regression	0.6966	0.7	0.72
Random Forest	0.7103	0.72	0.71
KNN Classifier	0.7034	0.7	0.7
Decision Tree	0.6345	0.64	0.63
Gradient Boosting	0.6690	0.72	0.67
SVC	0.6759	0.79	0.68
Hyper Logistic regression	0.7172	0.77	0.72
Hyper KNN	0.6483	0.65	0.65
Hyper Gradient Boosting	0.70345	0.71	0.7

With a precision of 0.7, an accuracy of 0.6966, a f1 score of 0.69 and a recall of 0.72, the logistic regression model likewise performed well. This shows that the logistic regression model was able to identify a patient's likelihood of having liver illness properly, however it could have been a little less accurate than the random forest classifier.

The K-Nearest Neighbors classifier had a f1 score of 0.7, an accuracy of 0.7034, precision of 0.7, recall of 0.7, and. This shows that the KNN classifier was able to identify a patient's liver disease status properly, however it may have performed somewhat worse than the random forest and logistic regression models.

With an precision of 0.64 , accuracy of 0.6345, a f1 score of 0.64 and recall of 0.63 the decision tree classifier performed the worst overall. As a result, it might not have been the ideal approach for this specific dataset to use the decision tree classifier.

The precision, accuracy, f1 score and recall of the gradient boosting classifier were 0.72, 0.6690, 0.68, and 0.67, respectively. This suggests that, although it may have been less accurate than the random forest and logistic regression models, the gradient boosting classifier was still able to correctly identify whether a patient had liver disease or not.

The precision, accuracy, f1 score and recall for the support vector classifier were 0.79, 0.6759, 0.68 and 0.68 respectively. This suggests that, compared to the random forest and logistic regression models, the support vector classifier may have been less balanced between accuracy and recall in its ability to identify whether a patient had liver disease or not

VI. Conclusion

Many machine learning algorithms were used to study the liver disease prediction dataset, and the findings indicate that random forest classifier and KNN may be the most appropriate models for this dataset, getting the greatest f1 score. With logistic regression and gradient boosting, hyperparameter adjustment can, nevertheless, result in better results.

Different machine learning models' performance can be affected by a variety of circumstances, therefore it's critical to thoroughly assess each model's performance and select the method that will be most effective for the given task. Other machine learning methods or feature selection strategies may

be used in future studies to increase the precision of liver disease prediction.

The investigation of the liver disease prediction dataset shows the positive results that may be obtained, and overall, the application of machine learning algorithms to medical diagnosis and prognosis has the potential to significantly enhance healthcare outcomes. To ensure accuracy and dependability, it is crucial to thoroughly evaluate the performance of such models before putting them into use in clinical settings.

REFERENCES

[1] Mohammed, A., Abou Neel, E. A., & Alnajjar, F. (2021). Machine Learning-based Prediction of Liver Disease Using Real-world Data from the National Health and Nutrition Examination Survey. Journal of Medical Systems, 45(4), 1-9. doi: 10.1007/s10916-021-01795-0

[2] Ali, A. M., Sakr, S., & El-Azab, D. S. (2020). Predicting Nonalcoholic Fatty Liver Disease using Real-world Data and Machine Learning. Journal of Medical Systems, 44(3), 1-10. doi: 10.1007/s10916-020-1531-8

[3] Wijarnpreecha, K., Thongprayoon, C., Panjawatanan, P., Ungprasert, P., & Cheungpasitporn, W. (2020). Machine Learning Models for Prediction of Hepatocellular Carcinoma Among Patients with Hepatitis B Virus Infection using Real-World Data. Clinical and Experimental Gastroenterology, 13, 411-419. doi: 10.2147/CEG.S260390

[4] Yang, J. D., Dai, J., Singal, A. G., Gopal, P., Addissie, B. D., Nguyen, M. H., & Wong, V. W. (2021). Machine Learning-based Prediction of Liver Fibrosis Progression using Real-World Data in Patients with Hepatitis C. Liver International, 41(4), 751-762. doi: 10.1111/liv.14716

[5] Li, Y., Lin, S., Guo, Y., Liu, S., Cheng, Y., & Zhou, Y. (2021). Machine Learning-based Prediction of Liver Fibrosis in Patients with Chronic Hepatitis B using Real-World Electronic Health Record Data. BMC Medical Informatics and Decision Making, 21(1), 1-9. doi: 10.1186/s12911-021-01507-4

[6] 6. Sharma, M., Nazar, C. M. J., & Wani, H. (2020). Prediction of liver fibrosis in non-alcoholic fatty liver disease using machine learning algorithms. Journal of medical systems, 44(12), 1-11.

[7] Wang, X., Chen, H., Cai, W., Liu, X., Chen, H., & Fang, Y. (2021). Machine learning-based prediction of liver cancer risk in patients with chronic hepatitis B. Frontiers in Oncology, 11, 642332.

[8] Song, Y., Cai, W., Chen, H., Fang, Y., Liu, X., Chen, H., & Li, S. (2021). Prediction of liver disease progression in patients with chronic hepatitis B using machine learning algorithms. BMC medical informatics and decision making, 21(1), 1-10.

[9] Zhang, J., Wang, Y., Zhang, X., Liu, J., Yang, Y., & Zhang, Y. (2021). Machine learning-based prediction of liver disease progression in patients with chronic hepatitis B. Journal of Medical Systems, 45(3), 1-9.

[10] Huang, C., Wang, X., Chen, H., Liu, X., Chen, H., Fang, Y., & Cai, W. (2021). Machine learning-based prediction of liver cancer risk in patients with chronic hepatitis B. Frontiers in Public Health, 9, 663271.

[11] Wang, X., Chen, H., Liu, X., Chen, H., & Fang, Y. (2020). Machine learning-based prediction of liver failure risk in patients with liver disease. Frontiers in Medicine, 7, 593. S. Vikrant, P. R. B, and B. H. S, "Policy for planned placement of sensor nodes in large scale wireless sensor network," KSII Trans. INTERNET Inf. Syst., vol. 10, no. 7, 2016, doi: 10.3837/tiis.2016.07.019

[12] V. Sharma, R. B. Patel, H. S. Bhadauria, and D. Prasad, "NADS: Neighbor Assisted Deployment Scheme for Optimal Placement of Sensor Nodes to Achieve Blanket Coverage in Wireless Sensor Network," Wirel. Pers. Commun., vol. 90, no. 4, pp. 1903–1933, Oct. 2016, doi: 10.1007/S11277-016-3430-6/METRICS.

[13] V. Sharma, R. B. Patel, H. S. Bhadauria, and D. Prasad, "Deployment schemes in wireless sensor network to achieve blanket coverage in large-scale open area: A review," Egypt. Informatics J., vol. 17, no. 1, pp. 45–56, Mar. 2016, doi: 10.1016/J.EIJ.2015.08.003.

[14] S. Vats and B. B. Sagar, "An independent time optimized hybrid infrastructure for big data analytics," https://doi.org/10.1142/S021798492050311X, vol. 34, no. 28, Jul. 2020, doi: 10.1142/S021798492050311X.

[15] S. Vats and B. B. Sagar, "Performance evaluation of K-means clustering on Hadoop infrastructure," https://doi.org/10.1080/09720529.2019.1692444, vol. 22, no. 8, pp. 1349–1363, Nov. 2020, doi: 10.1080/09720529.2019.1692444.